An Evaluation of Tennessee DOE's Internet Professional Development Initiative on the Attitudes and Performance of Teachers and Students

Final Report
Prepared by

Michael J. Hannafin
University of Georgia

June, 2002

DISCLAIMER

The summary presented in this report is the researcher's alone, and does not necessarily represent those of the Learning and Performance Support Laboratory, the University of Georgia or the State of Tennessee. The findings, recommendations, and conclusions are considered accurate and defensible. Extraordinary diligence was exercised to ensure the accuracy and integrity of this report. However, all source data used in the analysis were provided electronically by the Tennessee Department of Education and their contracting agents. Therefore, while the summaries, analyses, and interpretations based on these data are entirely those of the author, he is unable to warrant the accuracy of the source data. Where certainty could not be satisfactorily determined, data were excluded from a particular analysis or the study.

TABLE OF CONTENTS

Acknowledgements

The researcher wishes to acknowledge several current and former individuals in the Tennessee Department of Education for their assistance and cooperation in providing needed data and guidance: Jackie Shrago, Sue Goodwin, Jerry Bates, Ben Brown, and Soo-Hee Park. I also wish to acknowledge the SCR*TEC organization and associated individuals who hosted the on-line survey forms used by teachers and students throughout the study: Jim Nazworthy and Scott Finnerty. Finally, I acknowledge the efforts and assistance of colleagues at the University of Georgia: J.J. Hayden, Evan Glazer, and Janette Hill.

EXECUTIVE SUMMARY

This research was commissioned to examine the impact of TN-DOE's Connect-TEN professional development technology initiative. Participating teachers volunteered to complete the professional development program, implement a minimum of one Internet-based unit in their class, complete a succession of assessments related to their perceptions, and ensure that their students completed the same ratings and assessments. A total of 2350 volunteer teachers and their students completed an on-line, pre-unit survey prior to beginning their classroom implementation, rating their beliefs related to 8 teaching-learning constructs: Application, analysis, synthesis, usage, higher-order learning, motivation, transfer and relevance. The length of the Internet unit varied according to the complexity of the topic addressed and the discretion of the teacher, but most lasted 2-3 weeks. The same survey was administered immediately upon completion of the classroom Internet unit as well as prior to the end of the semester. The research study was implemented during Spring 2000.

Several questions were posed related to the influence of the Internet-based inservice program related to classroom teaching and learning among students and teachers in different geographic locations (urban, rural), different economic means (schools in wealthier, poorer economic areas), school level (elementary, middle, high school), and subject areas. The survey did not address specific technology knowledge or skills specifically, but rather the level of teaching and learning engendered and the nature of the activities associated with them following the completion of the state-sponsored professional development program. The purpose was to examine the extent to which everyday teaching practices shifted as a result of the classroom implementation to include more high-level activity rather than to assess perceptions towards technology per se.

Findings suggest that both teachers and students attribute performance changes to their Internet units. Ratings among both teachers and students tended to improve continuously from immediately prior to the study, to immediately following completion of a selected classroom unit, to the end of the study (near the end of the school year). Students universally rated more highly than teachers the frequency of activities related to improved, deeper thinking, with elementary school teachers generally indicating the least frequent implementation and high school teachers reporting the highest implementations. Student ratings were generally higher among urban schools and wealthier schools; however, all students reported progressively more favorable ratings throughout the study. Among students, elementary and high school urban students reported higher ratings than their rural counterparts; middle graders from both rural and urban settings reported similar ratings. Teacher ratings varied more as a function of school level, increasing from elementary, to middle, to high school.

However, actual performance improvements could not be confirmed using students' end-of-year test scores. The "gaps" selected by participating teachers for the Internet units implemented in the classroom could not be reliably classified per TCAP standards, and thus could not be reliably linked to the corresponding Terra Nova subtests. Therefore, only the more global subject area achievement indicator related to the Internet unit could be used. Predictably, student achievement on this indicator was not effected significantly.

Several implications can be drawn. First, considerable interest in activities considered to reflect higher-level teaching-learning processes were reported by both teachers and students—especially students. Perceptions of the use of higher-level, technology-enhanced teaching and learning were generally high across teachers and students. The progressive increase in the ratings by students suggests that the ratings improvements were sustained and increased and not simply due to the near-term novelty of the Internet classroom activity. Next, teacher reports of higher-level usage tended to increase from elementary, to middle, to high school. It seems possible that teachers perceive the curriculum at middle and high school as more amenable to technology and high-level activities, or that older students are seen as being more capable of such activities.

Two key issues could not be addressed satisfactorily but warrant further study: 1) the extent to which the inservice program and resulting classroom Internet unit improved student achievement; and 2) the nature and extent to which the classroom Internet units were implemented. Initially, the unit content was to be designated using TCAP gap identifiers, and matched with corresponding Terra Nova subtest scores. This was not possible, and the overall domain achievement indicators proved too broad to be of much value. It is important to address the impact of any initiative on student performance, using a range of useful, valid indicators including but not limited to standardized achievement measures. Since the classroom implementation of the Internet units was not observed, it was not possible to corroborate student or teacher ratings independently. It is important to provide such verification not only to corroborate ratings, but to identify key practices that account for differences in perceptions and performance. This should prove important in planning for and providing sustainable support to improve both research outcomes as well as classroom practices.

INTRODUCTION

This research was initiated at the request of the Tennessee Department of Education to examine the impact of its statewide professional development technology initiative. The focus was not on assessing the effect of, or beliefs about, technology per se, but the influence of technology-enhanced, Internet-based professional development on teaching-learning activities with or without technology. The initiative involved teachers receiving an Internet-based inservice program, and subsequently selecting one or more units to implement in their classrooms. The unit(s) selected represented a performance "gap", defined as TCAP area considered uniquely important, for which teachers developed and implemented Internet-based approaches within their classroom. Participating teachers volunteered to complete the professional development program, implement a minimum of one Internet-based unit in their class, complete a succession of assessments related to their perceptions, and ensure that their students completed the same ratings and assessments. Each teacher received nominal remuneration from TN-DOE for participation.

Goals & Background

This research was initially conceptualized to examine several questions related to both user perceptions and performance impact of the professional development initiative. The key questions are listed below:

- Do students and teachers attribute performance changes to Internet-based instruction?
- Do tools used for district, state and national evaluation purposes also provide useful information for teachers and students?
- Do the frequency and nature of Internet usage in the classroom influence student performance on standardized tests?
- Does Internet-based instruction influence learning for historically low achieving and/or economically disadvantaged students?
- Does Internet-based instruction influence the development of higher order thinking skills?
- Does Internet-based instruction influence the transfer of learning in one subject area to other subject areas?
- Does Internet-based instruction influence student transfer of knowledge and skill to everyday events?
- How do teachers use Internet in the classroom?

Initially, the research design involved a combination of on-line surveys administered through SCR*TEC using their web-based survey tool (Profiler), classroom observations of units as they were implemented, detailed interviews with select teachers and students, and standardized Terra Nova test data obtained routinely through the state's annual achievement assessment. In addition, use of the Value Added score was considered. Due to budget limitations, restrictions in the availability and use of the value added scores, and the lack of availability of observers, only two of these initial data sources were available: on-line ratings by participating teachers and students, and standardized test data for students in grades 3 through 8. This limited, to a significant extent, both the scope of the research and the capacity to address the research question. Given the considerable investment of both the TN-DOE staff and the participating teachers, however, the study proceeded with an aim to provide the best possible account of the professional development program's impact on students and teachers, and to identify limitations in the research accordingly.

Profiler Subscales

Based on the focus and intent of the project as identified by TN-DOE, several factors of interest were identified. An existing survey (Profiler) previously by TN-DOE, containing a total of 24 items, was adapted for use in the present study (See Appendix A). Based on an initial field-test, several items were re-worded and keyed to one or more of 8 subscales: 3 based on Bloom's Taxonomy (Application, Analysis, Synthesis), Higher-Level Learning (per Milken's criteria), Usage, Motivation, Transfer, and Relevance. Multi-item subscales were created to increase the reliability and validity of the survey in measuring the constructs being assessed. An initial field-test of the on-line survey was undertaken in Fall 1999 to establish the reliability and validity of the individual items as well as the discrimination performance of the items and the coherence and validity of the subscales.

Minor variations in wording were made to make the items relevant to both teachers and students and to maintain parallel content and focus. For example, the student version of one item, "I identify the major components of a topic to explain it to others," was phrased, "My students identify the major components of a topic to explain it to others." The survey was administered immediately prior to the target unit selected by the teacher, immediately upon completion of the unit, and near the end of the school year. The goal was to identify shifts in initial perceptions and to examine the durability of the shifts. K-12 teacher participants completed the survey, but due to the procedural complexities and widespread problems identified in the pilot test, only students in grades 3-12 completed the survey; K-2 students participated in the activities but did not complete the surveys. Where required, the teacher or aide assisted their students to minimize the effects of reading and limited computer testing familiarity.

Raters, students as well as teachers, rated each item based on the frequency with which they engaged in the activity specified using six Likert-type points that ranged from *Never* through *Daily*; these rating levels were assigned numeric weights that incremented from zero (for Never) to 5 (for Daily). Individual item ratings were tallied across the survey to yield individual scores for each subscale; the maximum scores for individual subscales ranged from 20-45. As total subscale scores approached the maximum possible score, the activities associated with the subscale were perceived as being more frequently evident; as the subscale scores approached zero, few-to-no associated activities were reported as being evident.

Bloom's Taxonomy: Application, Analysis, Synthesis. Three scales were defined as being consistent with the goals of the statewide inservice program. Nine application items, yielding a maximum subscale score of 45, measured the extent to knowledge and skills influenced activity in both everyday and formal education. An example application item was, "I use what I know about this topic in other school subjects." A total of seven analysis items totaling a maximum subscale score of 35 emphasized the ability to break down and reconstruct ideas and arguments logically. A representative analysis items was, "I decide which tasks are most important to solve a problem and put them in the right order." Synthesis items (5 total) focused on establishing connections among ideas and generating new interpretation based upon them. One synthesis item was, "When I learn something I combine with other things I know to create new ideas." Each was defined in order to examine the extent to which students and teachers perceptions related to them changed over the course of the study. The maximum synthesis score was 25.

Higher-Order Learning. This scale was defined based upon the criteria developed by the Milken's Foundation. It was included since the inservice program was designed specifically to promote higher-order learning through the implementation of activities designed to elevate the emphasis on non-rote and simple recall types of learning: Does this use of technology make it increasingly possible for the learner to engage in learning practices that lead to new ways of thinking, understanding, constructing knowledge and communicating results? A total of nine items comprised the scale yielding a maximum score of 45; an example item was, "I think of different ways to explain, report and demonstrate what I've learned."

Usage. A total of five usage items focused on how technology was used in the classroom—again as rated independently by teachers and by their students. The purpose was to benchmark the ways technology was being implemented prior to and subsequent to the inservice program, and to identify the extent to which students and teachers identified similar usage patterns. An example usage items was, "I use Internet in my class to talk to experts." The maximum possible score for the usage subscale was 25.

Motivation. Four motivation items, yielding a maximum score of 20, focused on the perceived influence of providing technology access on students' intrinsic motivation to learn, that is, their interest and willingness to engage learning for its own rewards rather than for teacher recognition or grades. A representative item was, "I stretch my thinking skills by taking on tasks that seem hard to do." Again, the focus for the study was on both the changes that occurred over time as well as the consistency between student and teacher ratings on motivation.

Transfer. Transfer items addressed the question of whether Internet-based instruction influenced student transfer of knowledge and skill to everyday events. A total of five items were included, yielding a maximum subscale score of 25; an example item was, "I connect what I learn in school to things I know that other people use in work or hobbies."

Relevance. A total of seven items (maximum subscale score of 35) were keyed to assess whether students use contemporary technology, communication networks and associated learning contexts to engage in relevant, real-life activity. The focus of this scale was not on transfer of specific knowledge or skills, but on the relevance to everyday life. A representative item was, "I can think of people's jobs that use information I learn in school."

PROCEDURES

Using the research questions as guideposts (Appendix B) and the TN-DOE Internet program as the defined intervention, several steps were undertaken. First, sampling requirements were determined based upon the requirements of the design and availability of project funds for participating teachers. Among the teachers and students statewide who completed the inservice program, target thresholds were set to stratify, to the extent possible, participation across three key demographic factors: geographic location (urban, rural), school economic means (based on federal definition of percentage of students eligible for free or reduced lunches), and school level (elementary, middle, high school). In addition, while not stratified, the subject area selected for the Internet unit was also identified for further analysis. While interest was expressed in both gender and ethnicity of teachers and students, TN-DOE staff advised against tracking or reporting data using these demographic factors.

The research study was implemented during Spring 2000. A total of 2350 volunteer teachers and their students were initially matched against the target participation requirements established for stratified sampling. It was presumed that, due to the nature of the study, a significant number of teachers and students would either drop from the study or fail to complete all aspects of the project. Therefore, a larger number of teachers and students were solicited initially; attrition was expected and assumed to be random across the project.

Each participating teacher and student completed the on-line pre-unit survey prior to beginning their classroom implementation. The survey, shown in Appendix A, contained 24 items, each requiring the raters to identify their beliefs related to technology in their selected Internet unit. Sample items included, "I use more than one source of information to form my ideas"; "I use Internet in my class to research topics," and so on. The survey solicited responses along a 6-point Likert-type scale, with corresponding descriptive labels (e.g., Never, less than every month, less than once each week, 1-2 times each week, 3-4 times each week, at least once each day). The length of the Internet unit varied according to the complexity of the topic addressed and the discretion of the teacher, but most lasted 2-3 weeks. The same survey was administered immediately upon completion of the classroom Internet unit as well as prior to the end of the semester.

Aggregation of Data

A total of 14 individual data sets were provided from three agencies. Sampling information and school demographic data for participating teachers was supplied by TN-DOE's ConnecTEN office; end-of-year TCAP-Terra Nova scores for participating elementary school students was provided by TN-DOE's Evaluation and Assessment Division; and on-line Profiler ratings for each phase by students and teachers were provided by SCR*TEC. Initial passes through the data revealed duplicate entries in the Profiler ratings supplied by students, and complete replication of teacher ratings across the pre-unit, post-unit and end of semester ratings. In addition, a number of students participated only in 1 or 2 of the assessments. Since this was anticipated in specifying the initial number of participants, only students who participated in each phase of the study were included. Given that teacher ratings were duplicated in their entirety, the end-of-semester ratings were used; consequently, no comparisons could be made regarding teachers' change of perceptions during the study.

NCE test scores were provided for students in grades 3-8. Since the above inclusion criteria ensured that students participated throughout the study, test scores were available for nearly all 3rd-through-8th graders, enabling the matching of unique test performance data with each student. However, the initial goal—to pair the Internet unit "gap" with the corresponding Terra Nova subscale score—proved impossible. The teacher-supplied unit descriptions were largely unlinkable to the corresponding TCAP standard. Therefore, Internet units were paired with the overall Terra Nova subject area scores rather than the more specific TCAP competency—a substantially broader comparison than initially planned.

The compilation of a common data set required the matching of identifying information across multiple data sets and file formats: district name and number, school level (grade), school name and number, percent of students eligible for free/reduced lunch, class number, teacher number, and student number. Some identifiers (district name and ID, school name and ID, school level, teacher name, subject selected for unit, percent eligible for free/reduced lunch) were provided by TN-DOE's ConnecTEN Office using the state designations. Profiler assigned student and teacher participants a number to track and link data across the study. TCAP-Terra Nova test scores were provided by participating class (with teacher and student last names). SQL was used to match/pair across data sets and compile required data from each into a common data set. Once pairings were made and the linked data verified for accuracy, all personally identifiable information for teachers and students was deleted.

Exclusion of Cases

Several factors were considered in the decision to include or exclude cases. First, where Profiler ratings for each phase were not provided, the cases were excluded for teachers and students. Where teacher and student ID were not matched, the case was excluded since within-person data could not be paired. Where it was not possible to associate student and teachers to the corresponding class ID, that is, student data could not be associated with a corresponding teacher, cases were excluded. In addition, in cases where multiple ratings were provided with the same date stamp by the same student or teacher, the cases were excluded since it was not possible to reliably identify the study phase. Finally, given the information provided, it was not possible to differentiate among students with the same last name within the same class; therefore, students in the same class with identical last names were excluded from the analysis. However, due to the nature of the study, occurring over an extended period of time and requiring multiple data sources for each, initial sampling requirements were set to account for attrition; the final sample was sufficiently large to permit the planned analyses

Efforts to Match TerraNova Scores to Specified Unit Gap

Initially, TN-DOE ConnecTEN staff attempted to classify the Internet unit gap descriptions identified by each teacher. Whereas the subject area for each Internet unit was readily identified, the correspondence between their TCAP gap description and the Terra Nova OPI subscales was not. Relatively few of the entries identified the TCAP gap specifically as requested; rather, the descriptions varied from a rationale for the unit, to evidence of the problem, to 1-2 work statements related to the subject. After an exhaustive classification effort by TN-DOE staff, roughly 10% were reliably classifiable. Given the number of subjects (8) and grades (K-12) for which units were implemented, the number of descriptions that could be linked to Terra Nova subscale scores was insufficient to permit reliable analysis. With the concurrence of TN-DOE, the corresponding subject area scores were used in lieu of the specific Terra Nova OPI score for the defined gap.

FINDINGS

Initial Screening and Data Reduction

More than 47,000 cases were contributed among the students and teachers who participated in at least one on-line survey phase of the study: pre-assessment, post-assessment, and end-of-year assessment. This figure also includes students and teachers who participated in multiple units. Since survey data could not be paired for comparisons without entries for each study phase, and since large numbers of students and teachers participated in all phases, only complete data sets were used. In addition, Terra Nova test scores were available for students in grades 3-8, which enabled the pairing of student ratings with performance on end-of-year assessments at these grades. Test performance data were not available for grades K-2 or 9-12 students. Appendices C-J contain detailed breakdowns of the means and standard deviations of student and teacher scores on all measures used in the study, along with summary data from the statistical analyses. In an effort to improve the readability of this report, descriptions and visual summaries of the analysis and findings are included in this report, and statistical summaries are mainly reported in the corresponding appendix.

Profile of Participants

Based on the above inclusion criteria, a total of 9667 students and teachers were included in the final analyses. Of the 8233 students, 2156 were in K-6 schools, 1940 in middle grades (7 through 9), and the remaining 4137 were high school students. Among the 1434 teachers, 856 taught at elementary schools, 406 in middle grades and 172 at high school. Participation was comparable for rural (52%) and urban (48%) schools; roughly 75% of the schools reported eligibility for free or reduced lunches, per federal criteria of a minimum of 49% student eligibility. These data are summarized in Table 1.

School Location	% Free-Reduced Lunch	School Level	Student	Teacher
Rural	< 50%	Elementary	476	202
		Middle	678	153
		High School	1793	74
	≥ 50%	Elementary	809	250
		Middle	353	46
		High School	149	4
Urban	< 50%	Elementary	397	161
		Middle	595	165
		High School	2184	93
	≥ 50%	Elementary	474	243
		Middle	314	42
		High School	11	1

Table 1. Student and teacher participant by geographic location, eligibility for free/reduced lunch, and school level.

Participation by Level, Free-Reduced Lunch, Geographic Location, and Subjects

Each participant was required to identify the subject area addressed in their selected Internet unit(s). As shown in Table 2, the content focus at both the elementary and middle school levels was generally in core academic areas: Language (including reading), Mathematics, Science and Social Studies. At the high school level, the content focus broadened as expected, but remained principally in the core academic areas.

Student Perceptions

The first set of analyses focused on student perceptions related to 3 areas defined within Bloom's Taxonomy (Application, Analysis, Synthesis). The remaining 5 areas defined prior to the study were: Higher-Order Learning, Motivation, Relevance, Transfer and Usage. The means and standard deviations and analysis source data for all student perception subscales are contained in Appendix C. Due to the large number of cases available and the resulting power of the analysis, relatively rigorous thresholds were established (p<.01) for testing differences statistically.

Level	Free Red.	Loc.		Ag	Arts	Bus & Voc	Fam & Cons	Health PE	Lang	Math	Fgn. Lang.	Sci	Social Stud	Total
Elem	< 50%	Rural	Student						104	149		111	112	476
			Teacher		15			4	97	44		28	14	202
		Urban	Student						22	122		40	213	397
			Teacher		10	1		5	43	54		24	24	161
	≥ 50%	Rural	Student						226	173		157	253	809
			Teacher		7	1		6	97	76		37	26	250
		Urban	Student						61	138		119	156	474
			Teacher		11	1		9	104	65		36	17	243
		Subtotal	Student						413	582		427	734	2156
			Teacher		43	3		24	341	239		125	81	856
Middle	< 50%	Rural	Student						86	147		109	336	678
			Teacher		2	8	4	3	51	35	1	24	25	153
		Urban	Student						142	139		115	199	595
			Teacher	1	11	14	3	16	46	34		21	19	165
	≥ 50%	Rural	Student						85	159		82	27	353
			Teacher		2	7	1	2	16	10		6	2	46
		Urban	Student						94	48		101	71	314
			Teacher		3	2		1	14	9		8	5	42
		Subtotal	Student						407	493		407	633	1940
			Teacher	1	18	31	8	22	127	88	1	59	51	406
HS	< 50%	Rural	Student		77	194	97	72	449	236	80	361	227	1793
			Teacher		3	13	2		23	5	4	18	6	74
		Urban	Student	14	92	281	52	106	546	253	166	354	320	2184
			Teacher		5	12	5	3	23	12	7	11	15	93
	≥ 50%	Rural	Student		12	18	14	24	18	20		43		149
			Teacher		1				2	1				4
		Urban	Student						11					11
			Teacher						1					1
		Subtotal	Student	14	181	493	163	202	1024	509	246	758	547	4137
			Teacher		9	25	7	3	49	18	11	29	21	172
Totals			Student	14	181	493	163	202	1844	1584	246	1592	1914	8233
			Teacher	1	70	59	15	49	517	345	12	213	153	1434

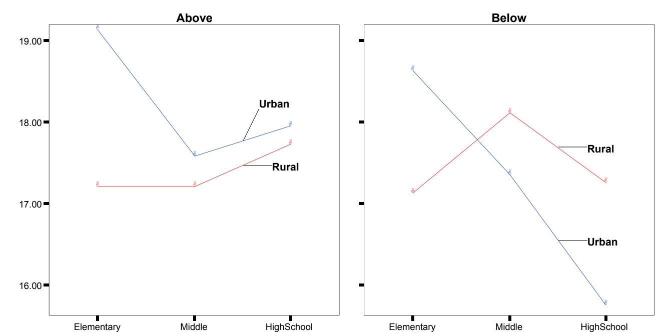
Table 2. Participation by subject area for which Internet unit was implemented.

Blooms Taxonomy Subscales

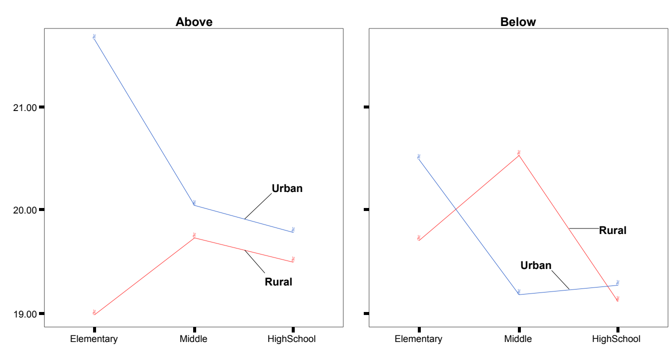
Significant overall effects were found on the Application, Analysis and Synthesis scales. Across the three subscales, student ratings improved significantly during the study, generally characterized by progressively and significantly higher ratings from pre, to post, to end of student assessments. This finding suggests that student perceptions improved following the Internet unit, and that this improvement was generally stable or increased from unit completion until the end of the school year; initial improvements in student perceptions were sustained. Significant MANOVA differences were found for economic means, school level, and geographic location, as well as the interactions between and among them. Effects for each scale were further refined using repeated measures ANOVAs for the three phases (pre, post unit, end-of-study) when the scales were administered (see MANOVA summary in Appendix C).

Student ratings on the each of the Bloom's Taxonomy subscales improved significantly over the study, generally becoming more favorable as the study progressed. The same patterns were evident for each of the subscales, with differences found for school level and economic means, but not for geographic location. However, some interactions involving the factors were detected. Generally, younger students (especially urban students) reported more favorable ratings for each scale than older students. Overall, the ratings of students from wealthier schools were higher as were the ratings of older students, and highest for younger students in wealthier schools. Interestingly, overall ratings did not differ for students in urban compared with rural schools.

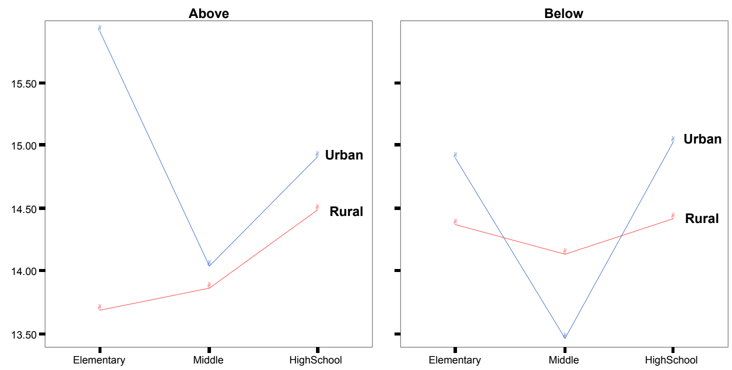
Application (max score=45)



Analysis (max score=35)



Synthesis (max score=25)

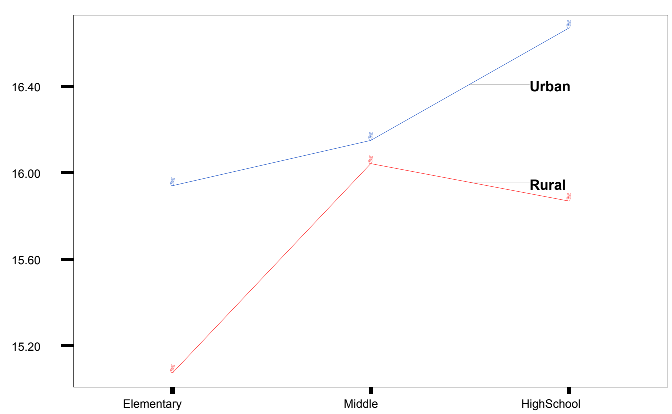


Usage

The Usage subscale focused on how technology usage was perceived by students. Usage ratings also increased over the course of the study. Ratings improved from pre-unit to post unit, suggesting an improvement in perceptions of the importance of technology use. The perceptions remained relatively stable from the post-unit to end-of-study assessment, suggesting that initial improvements in perceptions as a consequence of the Internet unit were sustained.

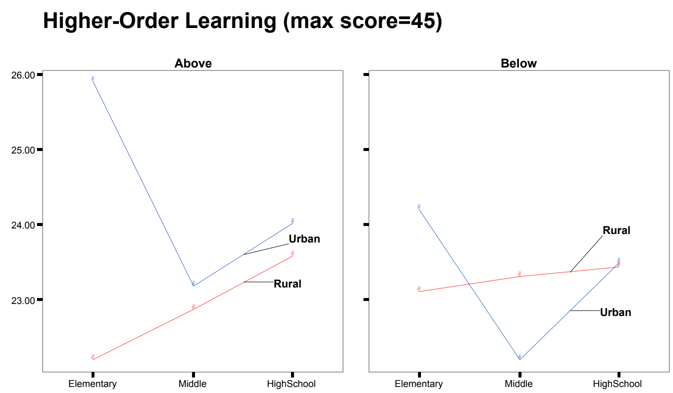
Unlike the ratings for the Bloom's Taxonomy subscales, significant MANOVA differences were found for school level and the interaction between school level and geographic location. Both older and younger urban students reported proportionately higher Usage ratings, with urban, rural middle school students reporting nearly identical ratings.

Usage (max score=25)



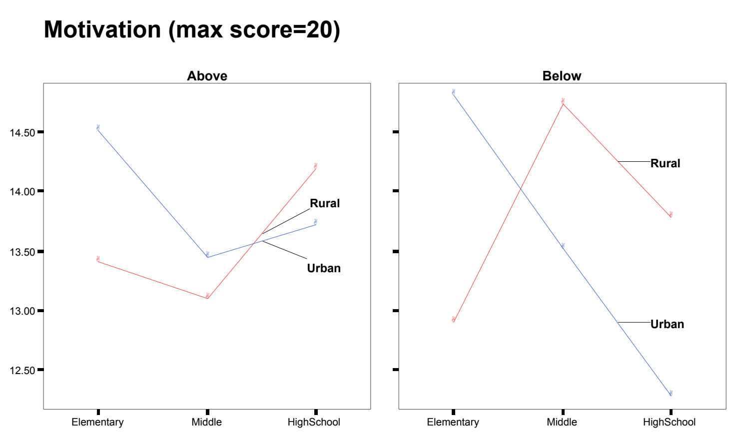
Higher-Order Learning

Ratings for higher-order learning changed over the course of the study, from pre-through-end-of-study assessments. Again, school level was significantly related to ratings for higher-order learning. However, two interactions were also detected. The first involved location and income differences, with wealthier urban students reporting proportionately higher ratings that wealthy rural students. Location and school level were also significantly related to ratings differences. As illustrated below, young urban students reported proportionately higher ratings than their rural counterparts; middle and high school students responded comparably. While geographic location was not, by itself, related to student ratings, significant interactions were found involving geographic location. Student perceptions of the role of technology on higher-order learning were influenced by the combination of geographic location, school level and economic means.



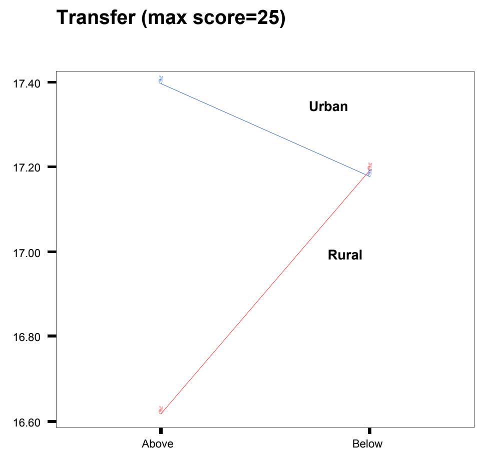
Motivation

All main effects and interaction combinations involving geographic location, school level, and economic means were significant for the motivation subscale. The main effect variance was mainly accounted for by the differences attributed to location and school level. In addition, an interaction between economic means and school level was found; students from lower income middle schools reported significantly higher ratings that their counterparts in higher income middle schools; elementary and high schoolers responded comparably within each level. Unlike the other scales, a significant interaction was also found among location, economic means, and school level. Again, the pattern for middle school raters from wealthier urban school was reversed from those in poorer rural schools, with lower income rural middle school students reporting high ratings while higher income urban students reported lower ratings.



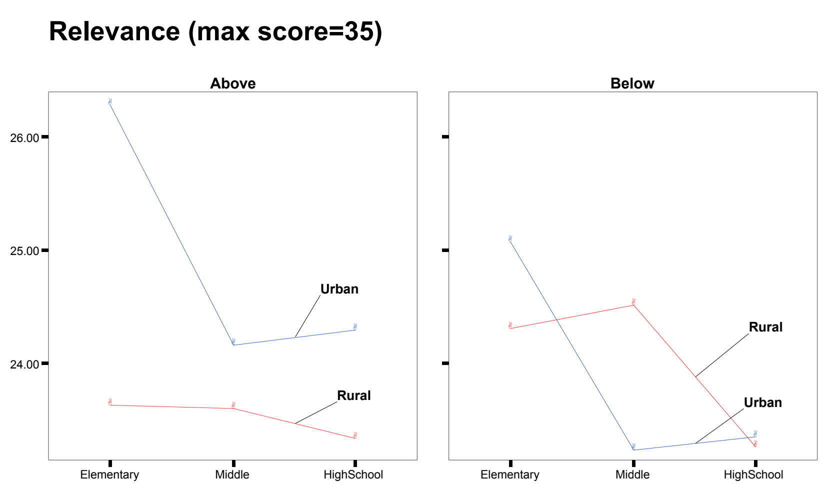
Transfer

As did the other scales, the transfer subscale scores changed significantly over time. This change, however, was influenced by geographic location. Middle school students rated transfer items comparably regardless of location, while elementary and high school students in urban location rated transfer far higher than did their peers in rural areas. An additional interaction was detected between economic means and geographic location. Students in lower income schools rated transfer items comparably to wealthier urban students, but wealthier rural students rated transfer items significantly lower.



Relevance

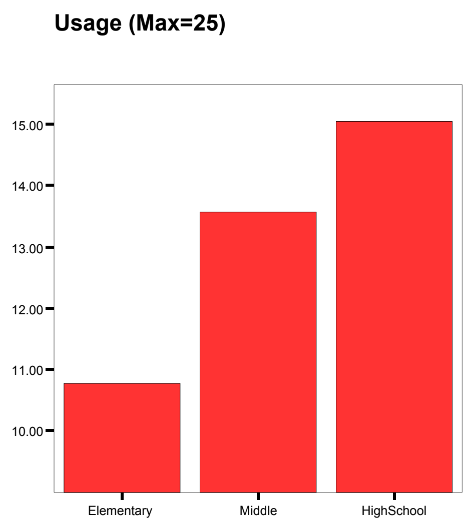
Perceived relevance scores changed significantly over the duration of the study, illustrating the same pattern as for the other subscales: higher post-unit scores that largely stabilized over time. The change in scores appeared to be influenced by school level. Again, elementary and high schoolers rated relevance proportionately higher regardless of location, while middle schoolers reported nearly identical ratings. The interaction among school level and geographic location was also significant. Students from wealthier and poorer urban schools rated comparably to rural students in poorer schools, while wealthier rural students provided the lowest ratings.

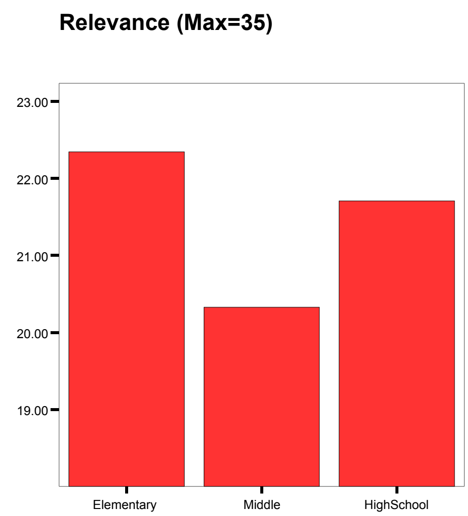


Teacher Perceptions

Whereas student perceptions were examined using the 3 survey assessments (before, after unit, end of study), only one assessment was available for teachers. Duplicate data appeared to be overwritten for each phase, yielding identical rating scores by each respondent across the 3 phases. Using the time-stamp as the point of reference, the responses were judged most likely the end-of-study responses; therefore, the analysis reflects teacher perceptions on the end-of-study assessment rather than repeated assessments throughout the study. As with the students, the first set of analyses focused on perceptions related to 3 areas defined within Bloom's Taxonomy (Application, Analysis, Synthesis), and the 5 areas assessed using the survey: Higher-Order Learning, Motivation, Relevance, Transfer and Usage. The means and standard deviations along with corresponding MANOVA summary data are contained in Appendix D. Again, relatively rigorous thresholds were established ($p<.01$) for testing differences statistically.

Unlike student ratings, overall differences were found only for school level; no other main effects or interactions. Among the subscales, school level differences were found only for usage and relevance; these are illustrated below using bar charts. High school teachers rated student usage highest, with middle school teachers next, followed by elementary teachers. A different pattern emerged for perceived relevance, with elementary and high school teachers ratings comparably but higher than middle school teachers.

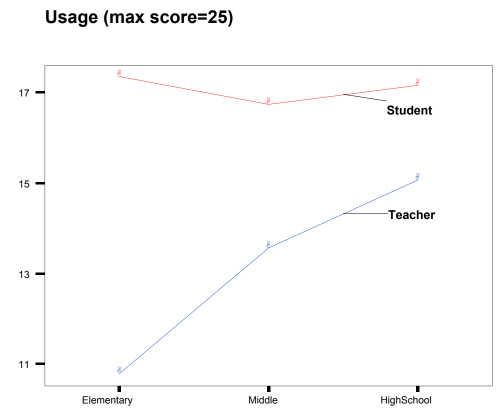




Perception Differences between Students & Teachers

Due to the aforementioned anomalies for teacher ratings, the differences between student and teacher ratings were examined statistically using MANOVA and follow-up ANOVA procedures where appropriate for the final ratings for each group. Mean and source data are shown in Appendix E.

Overall, significant differences were found between student and teacher ratings for each of the perception subscales, with students reporting higher, more favorable ratings than teachers across all scales. The interaction between school level and participant was also statistically significant. While student ratings were consistently higher than teacher ratings, teacher ratings increased progressively from elementary, to middle, to high school. Follow-up subscale ANOVAs indicated that the interaction was accounted for almost entirely by differences in rating for usage, with teacher ratings being progressive higher from elementary, to middle, to high school but consistent students.



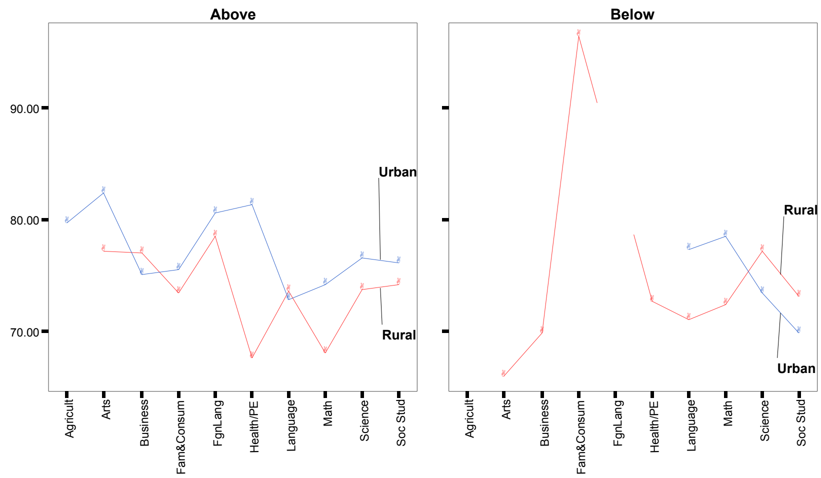
Ratings Differences by Subject Area

One of the analyses of interest focused on differences in perceptions by subject area for which the Internet unit was deployed. Since teacher ratings were duplicates, no difference scores could be calculated and no rating pre-to-post unit difference changes among teachers or between teacher and student ratings could be made. To examine differences between student and teacher ratings by subject of unit, the final ratings were used. Summaries and source data for students and teacher ratings by subject area are shown in Appendix F and Appendix G respectively.

Subscale Ratings by Subject Area

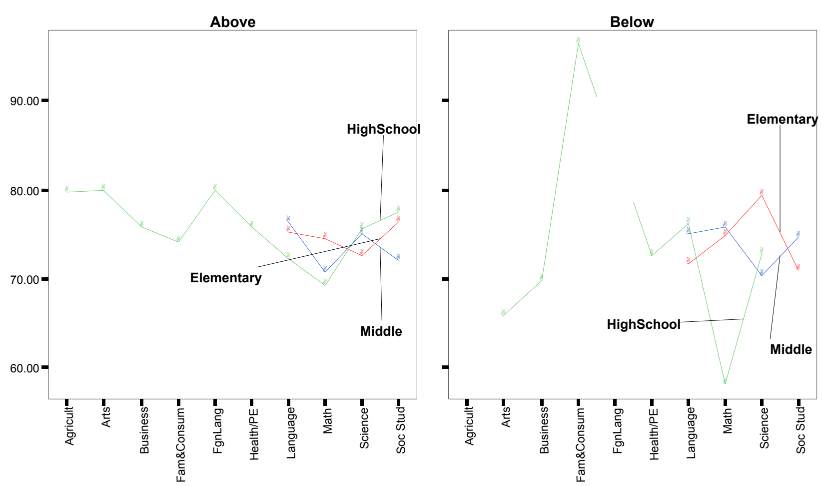
Overall, significant differences were found across all subscale ratings when organized by Internet unit subjects. For students, ratings were also influenced by the interplay among subject, geographic location and the economic means of the school. This interaction is typified below, using the overall survey scores, since the pattern was virtually identical across subscales. In schools of greater and lower economic means, urban and rural students consistently reported similar or higher ratings across subjects, but rural students in family and consumer sciences classes in lower income schools reported significantly higher ratings than any other groups—a pattern consistent across subscales.

Overall Survey Score



As shown in the following figure, this effect was influenced mainly by high school student ratings, as no middle or elementary students in the study were completing family and consumer education courses. In addition, high school students in lower income schools mathematics classes reported less evidence of the project activities (per the inservice unit and scale focus) than other high school classes as well as middle or elementary math students.

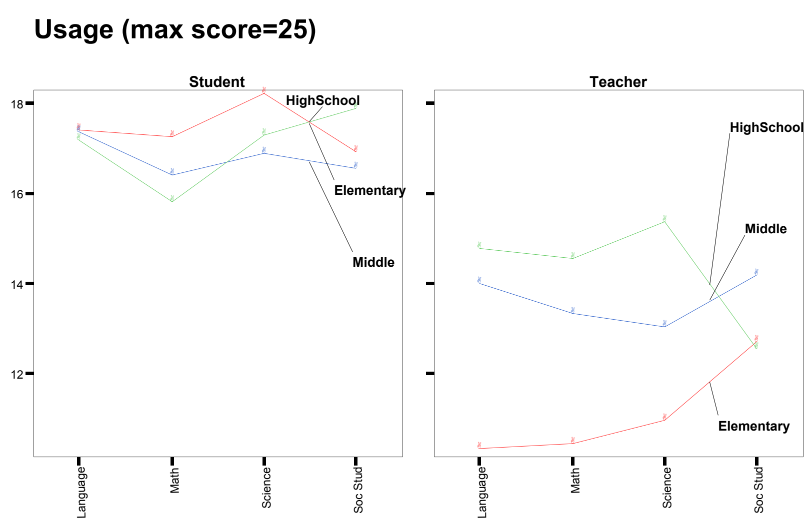
Overall Survey Score



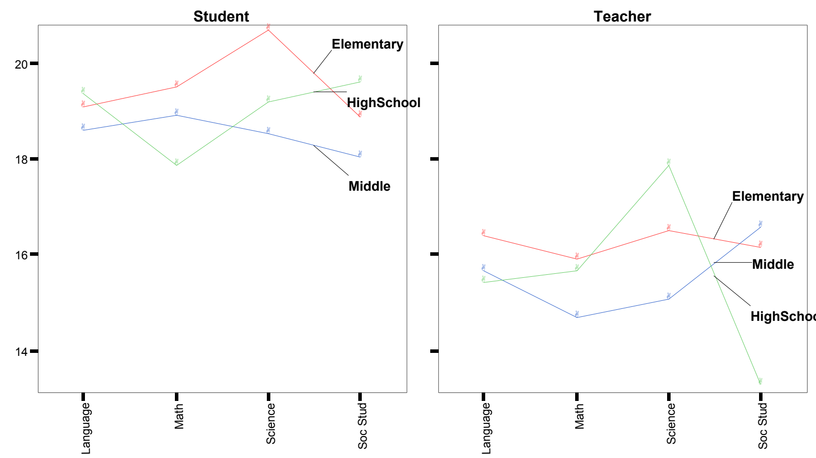
Teacher ratings were generally consistent across subject and level, as evident by the scores across subject and school level (note that low- and non-participation levels appear to have a disproportionate effect on the graphic representations, but not sufficient for reliable statistical differences). No subscale rating patterns varied reliably by the different subject areas or the interplay between subject area and the other factors included in the study.

Differences between Student and Teacher Ratings

The relationship between student and teacher ratings on each Profiler subscale is presented in Appendix H and illustrated in Appendix I. On further inspection, since several subject areas were only available at the high school level, an additional analysis of rating differences was undertaken in which several high school only courses were excluded. This had the effect of equating student and teacher ratings across those subjects where Internet units were actually taught. The summary, along with graphical summaries for all subscales by subject and participant, is contained in Appendix J. For the set of common subjects, including those considered "core" academic subjects (i.e., math, science, language, social studies), significant overall effects were found for Internet unit subject area and the interaction between subject, participant, and school level. Follow-up ANOVAs for individual subscales indicated significant interaction effects for the application and usage scales. This interaction is illustrated below. As usual, students reported significantly higher and consistently similar ratings across subject areas for both subscales, while the pattern of teacher ratings varied both from subject-to-subject and from subscale-to-subscale. High school teachers tended to report higher usage ratings than elementary or middle school teachers except for social studies where they reported the lowest ratings, but their application ratings were similar to elementary and middle school teachers. Elementary school teachers consistently reported the lowest usage among the school levels, but had the highest application ratings overall.



Application (max score=45)



Relationships between Perception & Performance Measures

One of the primary interests in this research was to examine the relationships between and among perceptions and student performance in the subject areas of their Internet unit. Initially, TCAP student data related to the specific "gaps" addressed in the Internet units was to be used as the criterion measure, which was considered to represent a reasonably close link between the content coverage defined in the unit selected and statewide assessment of knowledge and skill development. However, teacher description of the Internet unit failed to identify their gap with the corresponding Terra Nova test scores sufficiently to permit their use. Therefore, Internet units were linked with the corresponding overall subject scores for each student in grades 3-8. NCE scores for each student were obtained for the subject area identified in the Internet unit description. This was less-than-optimal since, to the extent the Internet units improved student achievement, it was most likely to be evident on tests of the knowledge and skills addressed in that unit. Lacking "close" links between the unit coverage and the assessment items, differences were unlikely to emerge or simply be subsumed under the more inclusive scope of the subject area tests. Test score data were not available for high school students.

Regression analyses were undertaken to identify whether subscale ratings were significantly related to NCE scores. As expected, none of the subscale ratings proved a significant predictor of student NCE performance in the corresponding Internet unit subject area. Among the variables studied, only the eligibility for free/reduced lunches proved a reliable predictor, with students in higher-income schools performing consistently higher on the achievement scores than students in lower-income schools. Interestingly, geographic location (urban, rural) was not a reliable predictor of student achievement, suggesting that while school income mattered, geographic location did not.

Summary by Question

The following summarizes findings by the specific research question initially posed represented as an over-arching question with a series of sub-questions.

How does Internet based instruction impact teaching and learning?

Do students and teachers attribute performance changes to Internet-based instruction? Survey evidence suggests that both teachers and student attribute performance changes—involving the teacher and the student—to their Internet units. However, apart from changes in perceptions, performance improvements were not found using students' end-of-year test scores. Since observation of teacher implementation of the Internet units was not undertaken, teacher performance changes could not be assessed.

Do tools used for district, state and national evaluation purposes also provide useful information for teachers and students? This question could not be assessed in the present study, since planned observation and interviews by TN-DOE personnel were not possible.

Does the frequency and nature of Internet usage in the classroom influence student performance on standardized tests? No evidence was found to establish a causal link between Internet usage, reported by either teacher or student, and student performance on corresponding subject areas on end-of-year tests.

Does Internet-based instruction influence learning for historically low achieving and/or economically disadvantaged students? In most cases, students in less affluent urban and rural school rated learning activities comparably to their peers in more affluent schools, suggesting that no compensatory or unique effect for the specific target group. While overall the ratings were comparable across more and less affluent schools, a few instances were found where ratings were differentiated by geographic location and/or economic means. Student ratings were generally higher among urban schools and wealthier schools. However, all students including this target group reported progressively more favorable ratings throughout the study.

Does Internet-based instruction influence the development of higher order thinking skills? Results on the Higher-Order Thinking Skills subscale used in this study indicated that both teachers and students perceive improvements in higher-order thinking skills. Again, however, corroboration in the form of student test performance was not evident.

Does Internet-based instruction influence the transfer of learning in one subject area to other subject areas? Does Internet-based instruction influence student transfer of knowledge and skill to everyday events? Results on the Transfer as well as the Relevance subscales indicated perceived improvements as reported by both students and teachers.

How do teachers use Internet in the classroom? The usage subscale revealed both similarities to and differences from the other subscales as reported by teachers and students. Among students, elementary and high school urban students reported higher ratings than their rural counterparts; middle graders from both rural and urban settings reported similar ratings. Teacher ratings varied more as a function of school level, with usage ratings increasing from elementary, to middle, to high school.

Can an Internet-based project designed for one State be transplanted to or adapted by another State? This question was conceptualized initially as a scale-up of the Tennessee program to other interested states. It was not, however, addressed in this study.

Does Internet based instruction influence the development of higher order thinking skills, relevancy and motivation as defined by Milken?

Does this use of technology make it increasingly possible for the learner to engage in learning practices that lead to new ways of thinking, understanding, constructing knowledge and communicating results? Both teachers and students rated items keyed to Milken's definition of higher-order learning favorably. Since the initial goal of linking specific Terra Nova subtest scores to selected gaps was not possible, it was not possible to equate the nature of the learning from either the Internet unit or achievement items as higher-order.

Are learners using contemporary technology, communication networks and associated learning contexts to engage in relevant, real-life applications of academic concepts? Again, both teachers and students rated items related to the relevance subscale favorably. Both urban and rural elementary and high school students rated relevance items high (especially at the elementary level), but middle school students responded comparably regardless of geographic location.

Is quality access to technology and telecommunications increasing the intrinsic motivation of learners to learn? Items keyed to the motivation subscale were generally rated positively by both students and teachers, with students reporting higher ratings than teachers; ratings improved steadily during the study. Among students, those in less affluent middle schools rated motivation items quite high compared to those in more affluent schools, while the pattern was reversed at elementary and high schools where more affluent students provided higher ratings. Among teachers, motivation ratings were not statistically different across geographic location or economic means.

Does Internet-based instruction influence the development of higher order thinking skills as identified in Bloom's taxonomy?

Student and teacher responses to the Application, Analysis and Synthesis subscale items were uniformly positive, becoming increasingly so among teachers at the middle, then high school levels compared with the elementary school level. Among students, the ratings were progressively higher at each phase of the study, suggesting a continued improvement in perception. Only urban elementary school student ratings on each scale remained roughly comparable across the study, but their ratings were consistently the highest among the three levels.

Implications, Issues & Recommendations

Numerous issues arose during this research, which are divided into four primary types: design, implementation, analysis and interpretation. With regard to design, the over-arching goal—to gather empirical data related to program implementation as it relates to student and teacher performance—is important and very much in keeping with current and future accountability requirements. To be most useful to all concerned, it is important to account for design and measurement requirements early in the process to test and validate both the procedures and the measures as well as to ensure that the approach will yield the data required and adequately address the questions and concerns. In this case, the program implementation was already underway, preemptively influencing the questions that could be posed, the method of implementation, and the data and time available. In effect, the current design addressed the question, “What kinds of data were available?” rather than “What kinds of data are needed?” In some cases, the research focused on what could be addressed rather than what should be addressed. Ideally, the evaluation-research design should be developed concurrently with the program to optimize the benefits of each.

Next, implementation of the research was hampered by limited resources and distributed responsibilities. Several individuals and organizations participated in various aspects of the study, and were exceptionally helpful, but all were already over-extended in meeting their existing responsibilities. The data sources identified for the study were “owned” or generated by numerous individuals and agencies, causing delays in identifying the appropriate agent, gaining authorization and approval, and eventually negotiating the release of the requested information. In some cases, data were not collected, such as observations of classroom implementations. In others, incorrect data were sent causing delays and wasted effort for all concerned. Two significant limitations in the study resulted from the inability to equate Internet units with TCAP gaps and the inability to corroborate ratings—both implementation problems. Investments in time, teachers and technology are very significant, and warrant a serious, sustained and dedicated commitment to program implementation. In this regard, it would likely have been sufficient to sample a smaller subset of the study’s population and commit to a more deliberate implementation to ensure that the needed data were available.

Third, a number of issues arose during data analysis. Numerous records (teacher, student) were excluded due to duplicate identification data. This was most problematic for teacher ratings, which were duplicated or over-written with the same entries during the entire study. It was not possible, as a result, to examine changes in teacher ratings during the study—an analysis that was available for students and provided interesting data on sustained changes in perception. Whereas Terra Nova subtest scores, keyed to the selected Internet unit gaps identified by each teacher, were available, the teacher descriptions could not be equated. Consequently, an even broader indicator of domain achievement (overall subject area score) served as the achievement indicator—a poor indicator of the impact of the selected unit. However, a larger issue needs to be addressed. Even if the gaps and subtest scores had been equated, they would still have been fairly global estimates of achievement related to the Internet-based unit—relatively few of the items addressed in the test would likely have been covered in the units. To the extent it is important to determine what has been learned from a given unit, more unit-specific assessments are needed. These may be weighed with, or against, more global achievement indicators, but they are likely to be far better estimates of learning in a given unit. The issue of which indicators are to be taken as evidence of impact on student learning is important and needs to be weighed carefully; clearly, considerable learning of different types has occurred. Based on the generally favorable ratings reported across the subscales, it seems that both teacher and students perceive important roles for technology in everyday teaching and learning. Ideally, some of the payoff of this interest and role is in measured achievement; a good deal of benefit, however, may also lie elsewhere such as improved motivation, attendance, persistence, independence, and so on. The quest for achievement indicators is important, but needs to more closely address program implementations and needs to be viewed more inclusively with other indicators.

Finally, the interpretations provided in this report cannot provide equal attention to statistical differences in the data and the practical importance of those differences. Given the large sample, it is possible to detect reliable statistical differences that are of little practical importance, and vice-versa. The reporting of statistical differences avoids the temptation to over-interpret group findings, effectively raising the standard required to describe an apparent difference as “real” or “reliable.” Given the large number of factors and variables included in the analyses, it is often wise to adopt conservative interpretation standards, as done in this report. To a significant extent, the interpretations have been limited by the absence of corroborating data. Observations, for example, might have shed light on usage pattern shifts among teachers from elementary, to middle, to high school; one can speculate in a general sense about such findings in this study, but observational data might have provided specific explanatory evidence. Likewise, while no evidence was found linking the professional development program and classroom implementation to student achievement, the lack of closer criterion-based measures rendered this goal futile. To the extent student achievement was influenced in the present study, the available measures were not adequate to detect it. Interpretation of these findings, therefore, is necessarily minimal.

Limitations

Several factors influenced both the implementation of the inservice program, the conduct of the study, and the analysis and summaries contained in this report. In some cases, the constraints are significant and altered much of what could be implemented and reported; in others, they were less fundamental but affected the process nonetheless.

First, the study involved assessing a process and implementation that was already underway rather than a design to optimize implementation and assessment. Many of the activities, data gathering instruments and systems, and reporting mechanisms had already been identified. The task for the study was to overlay an approach that best utilized those approaches and decisions, and to minimize the additional administrative and reporting burdens to the teachers and students. While this likely made the project seamless and transparent at the school level, a variety of methods and measures of potentially great use were not available.

Perhaps the most significant constraint was the inability to obtain accurate, reliable identification of Internet unit “gaps” that could be subsequently linked to specific Terra Nova test reports. Much of the initial design of the study was predicated on linking perceptions and implementations to widely recognized student performance measures. That is, teachers were to identify specific TCAP gaps based on state curriculum priorities, as well as to create and implement Internet-based instruction in their classes designed to address the selected gaps. End-of-year test data was then to be keyed to the specific subtest related to the identified gaps—an approach designed to provide student performance data on the very skill and knowledge for which the Internet units were designed. Since the unit descriptions were unclassifiable, only the more global achievement subject area scores could be used, which included a great deal more coverage than the Internet units. Consequently, the reported lack of a significant impact on achievement scores may understate the impact of the initiative; it was not possible to conduct the intended analysis.

Next, the survey items used in this study were examined during a field test to determine scale factors, validate content, and establish reliability. A number of revisions were recommended both in the number as well as content of the questions. However, the revised items were not made available via Profiler, and the problem was not detected until well into the current study. A decision was made to continue with the scale in use for the duration of the study, since examination of shifts in rater responses to items and subscales were critical to the design. Additional factor analyses and item validation was done upon completion of the study and subscales re-keyed accordingly. As predicted, a high degree of intercorrelation was found among subscale scores, in part due to multi-keyed items, in part due to similarities among the concepts, and in part due to item construction problems that could have been avoided if revised versions were employed. The survey is considered adequate for purposes of this study, but not optimal for differentiating among the factors that drove its design.

Next, in addition to the problem related to the usability of teacher unit descriptions, several planned data sources were not available. The individuals initially designated by TN-DOE to observe Internet unit implementations and interview students and teachers were unavailable due to budget and time constraints. As a result, data planned to corroborate teacher and student ratings, as well as to characterize the implementations, was likewise unavailable. The inability to corroborate either ratings or implementations limits the extent to which the trustworthiness of the data can be verified. It was assumed that implementations occurred as planned, and ratings were provided by corresponding students and teachers, but this could not be verified. Indeed, numerous cases of duplicate data and/or identification numbers were detected; such data were routinely excluded.

As with any large-scale, field-based research, it is impossible to verify all aspects and ensure the integrity of all accounts and data. However, the teachers and students are believed to have participated in an honest and responsible way. There was little to be gained in subverting the research; indeed the level of participation was so substantial as to render isolated cases of no consequence. In all cases of uncertainty, data were excluded from analysis. The above limitations, while potentially serious in some cases, reflect more on the integrity of the research methods than on the participating teachers and students. The findings of this study may or may not have been different from those reported; the limitations cited are provided to frame and qualify the findings, not to summarily dismiss them.